Jensen divergence based on Fisher's information

Pablo Sánchez-Moreno, Alejandro Zarzo and Jesús S. Dehesa

***Abstract*—The measure of Jensen-Fisher divergence between probability distributions is introduced and its theoretical grounds set up. This quantity, in contrast to the remaining Jensen divergences, is very sensitive to the fluctuations of the probability distributions because it is controlled by the (local) Fisher information, which is a gradient functional of the distribution. So, it is appropriate and informative when studying the similarity of distributions, mainly for those having oscillatory character. The new Jensen-Fisher divergence shares with the Jensen-Shannon divergence the following properties: non-negativity, additivity when applied to an arbitrary number of probability densities, symmetry under exchange of these densities, vanishing if and only if all the densities are equal, and definiteness even when these densities present non-common zeros. Moreover, the Jensen-Fisher divergence is shown to be expressed in terms of the relative Fisher information as the Jensen-Shannon divergence does in terms of the Kullback-Leibler or relative Shannon entropy. Finally the Jensen-Shannon and Jensen-Fisher divergences are compared for the following three large, non-trivial and qualitatively different families of probability distributions: the sinusoidal, generalized gamma-like and Rakhmanov-Hermite distributions.**

***Index Terms*—Jensen divergences, dissimilarity measures, discrimination information, Shannon entropy, Fisher information.**

I. Introduction

The study of the measures of similarity between probability densities is a fundamental topic in probability theory and statistics *per se* and because of its numerous applications and usefulness in a wide variety of scientific fields, including statistical physics, quantum chemistry, sequence analysis, pattern recognition, diversity, homology, neural networks, computational linguistics, bioinformatics and genomics, atomic and molecular physics and quantum information. The most popular measure of similarity between two probability densities $\rho_1(x)$ and $\rho_2(x)$ is possibly the Jensen-Shannon divergence [1], [2], which is defined as

$$JSD[\rho_1,\rho_2] = S\left[\frac{\rho_1+\rho_2}{2}\right] - \frac{S[\rho_1]+S[\rho_2]}{2}, \qquad (1)$$

where $S[\rho]$ denotes the Shannon entropy of the density $\rho(x)$, $x \in \Delta \subset \mathbb{R}$, given by

$$S[\rho] = -\int_\Delta \rho(x)\ln\rho(x)dx.$$

According to Eq. (1), the Jensen-Shannon divergence quantifies the Shannon entropy excess of a couple of distributions with respect to the mixture of their respective entropies. It can also be expressed as

$$JSD[\rho_1,\rho_2] = KL\left[\rho_1, \frac{\rho_1+\rho_2}{2}\right] + KL\left[\rho_2, \frac{\rho_1+\rho_2}{2}\right],$$

indicating that the Jensen-Shannon divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence (KLD in short) or relative Shannon entropy (also called Kullback divergence) defined [3], [4] by

$$KL[\rho_1,\rho_2] = \int_\Delta \rho_1(x)\ln\frac{\rho_1(x)}{\rho_2(x)}dx.$$

The Jensen-Shannon divergence as well as the KLD are non-negative and vanish if and only if the two densities are equal almost everywhere. Unlike the KLD, the Jensen-Shannon divergence has two additional important characteristics: it is always well defined (in the sense that it can be evaluated even when ρ_1 is not absolutely continuous with respect to ρ_2), and its square root verifies the triangle inequality so that the square root of $JSD[\rho_1,\rho_2]$ is a true metric in the space of probability distributions [5]. Furthermore, it admits the generalization to several probability distributions [1] in the following sense: let be a vector $\boldsymbol{\omega} = (\omega_1,\omega_2,\ldots,\omega_N)$ and a set of N probability densities $\{\rho_j(x)\}_{j=1}^N$; the Jensen-Shannon divergence among these probability densities is given by

$$\begin{aligned} JSD_{\boldsymbol{\omega}}[\rho_1,\ldots,\rho_N] = S[\omega_1\rho_1+\cdots+\omega_N\rho_N] \\ -\omega_1 S[\rho_1]-\cdots-\omega_N S[\rho_N], \end{aligned}$$

where the nonnegative numbers $\omega_i > 0$, for $i = 1,\ldots,N$, such that $\sum_i^N \omega_i = 1$, are weights properly chosen to indicate the relative relevance of each density. This is very useful for certain applications such as in bioinformatics, diversity and atomic physics where there are situations in which it is necessary to measure the overall differences of more than two probability distributions. Notice that for $N = 2$ one has

$$JSD_{\boldsymbol{\omega}}[\rho_1,\rho_2] = S[\omega_1\rho_1+\omega_2\rho_2]-\omega_1 S[\rho_1]-\omega_2 S[\rho_2],$$

so that it simplifies to the expression (1) in the case $\omega_1 = \omega_2 = \frac{1}{2}$.

This divergence has been extensively applied in numerous literary, scientific and technological areas ranging from information theory [1], [6], [7], statistical and quantum mechanics [8] to bioinformatics and genomics [9], [10], atomic physics [11]–[14] and quantum information [15], [16]. Let us just mention that it has been used as a tool to study EEG records [17], to segment symbolic sequences [18], to measure the complexity of genomic sequences [9], [10], to analyze literary texts and musical score [19], to quantify quantum phenomena such as entanglement and decoherence [15], [16] and to understand the complex organization and shell-filling patterns

P. Sánchez-Moreno is with the Department of Applied Mathematics and the Institute Carlos I for Theoretical and Computational Physics, University of Granada, Granada, Spain

A. Zarzo is with Department of Applied Mathematics, Polytechnic University of Madrid, Madrid, Spain, and the Institute Carlos I for Theoretical and Computational Physics, University of Granada, Granada, Spain

J.S. Dehesa is with the Department of Atomic, Molecular and Nuclear Physics and the Institute Carlos I for Theoretical and Computational Physics, University of Granada, Granada, Spain

of the many-electron systems all over the periodic table of chemical elements [11]–[14].

Nevertheless, the Jensen-Shannon divergence is, at times, weakly informative or even uninformative, mainly because it depends on a quantity of global character (the Shannon entropy) in the sense that it is hardly sensitive to the local fluctuations or irregularities of the probability densities. So, by definition, this divergence has serious defects to compare probability densities with highly oscillatory character. This is often the common situation in many fields, such as e.g. in the quantum-mechanical description of natural phenomena. To illustrate it, let us consider the simple case of the motion of a particle-in-a-box (i.e., in the infinite well $V(x) = 0$, for $0 < x < 1$, and $+\infty$ elsewhere) [20]. The stationary states of the particle are characterized by the sinusoidal probability densities

$$\rho_n(x) = 2\sin^2(\pi n x); x \in (0,1), \tag{2}$$

and $\rho_n(x) = 0$ when $x \notin (0,1)$, where $n = 1, 2, \ldots$ indicates the energetic level and label of the state. The divergence between the n-th quantum state $\rho_n(x)$ and the ground state $\rho_1(x)$ is studied in Figure 1 by means of the Jensen-Shannon measure $JSD[\rho_n, \rho_1]$. We observe that this divergence tends rapidly to a constant, so that it is not informative enough about the enormous differences between these two probability densities.

The case of a particle-in-a-box and other cases pointed out later show the necessity for defining a new divergence to be able to measure the similarity between two or more oscillating probability densities in a much more appropriate quantitative form. This is the purpose of our work: to introduce the Jensen-Fisher divergence, which depends on an information-theoretic quantity (the Fisher information [21], [22]) with a locality property: it is very sensitive to fluctuations of the density because it is a gradient functional of it. This is done in Section II, where the definition of the new divergence is given and its main properties are shown. Then, in Section III the Jensen-Shannon and Jensen-Fisher divergences are compared in the framework of an information theoretic plane for various cases properly chosen to illustrate the relative advantages and disadvantages of these two quantities; namely, the sinusoidal, generalized gamma and Rakhmanov-Hermite probability distributions. Finally, some conclusions and open problems are given.

II. The Jensen-Fisher divergence measure

In this Section we define a new Jensen divergence between probability distributions based on the Fisher informations of these distributions, and we study its main properties. In doing so, we follow a line of research similar to that of Lin [1] to derive the Jensen-Shannon divergence.

Let X be a continuous random variable with probability density $\rho(x)$, $x \in \Delta \subset \mathbb{R}$. The (translationally invariant) Fisher information of $\rho(x)$ is given [21], [22] by

$$F[\rho] = \int_\Delta \rho(x) \left[\frac{d}{dx}\ln\rho(x)\right]^2 dx, \tag{3}$$

and the relative Fisher information between the probability densities $\rho_1(x)$ and $\rho_2(x)$ is defined [23] by the directed divergence

$$F_{\rm rel}[\rho_1,\rho_2] = \int_\Delta \rho_1(x)\left[\frac{d}{dx}\ln\frac{\rho_1(x)}{\rho_2(x)}\right]^2 dx.$$

It is known that this quantity is non-negative and additive but non symmetric. The relative symmetric measure defined by

$$\begin{aligned} G[\rho_1,\rho_2] &= F_{\rm rel}[\rho_1,\rho_2] + F_{\rm rel}[\rho_2,\rho_1] \\ &= \int_\Delta (\rho_1(x)+\rho_2(x))\left[\frac{d}{dx}\ln\frac{\rho_1(x)}{\rho_2(x)}\right]^2 dx, \end{aligned}$$

is called Fisher divergence [23], which has been recently used in some applications to study the complexity and shell organization of the atomic systems along the Periodic Table [12], [14]. As in the Shannon case, this divergence is nonnegative and it vanishes if and only if $\rho_1(x) = \rho_2(x)$ for any $x \in \Delta$, but it is undefined unless that $\rho_1(x)$ and $\rho_2(x)$ be absolutely continuous with respect to each other.

To overcome these problems of the $F_{\rm rel}$ and G divergences, we define a new directed divergence between the probability densities $\rho_1(x)$ and $\rho_2(x)$ as

$$\overline{F}_{\rm rel}[\rho_1,\rho_2] = \int_\Delta \rho_1(x)\left(\frac{d}{dx}\ln\frac{\rho_1(x)}{\frac{\rho_1(x)+\rho_2(x))}{2}}\right)^2 dx.$$

This quantity is nonnegative because it vanishes if and only if $\rho_1(x) = \rho_2(x)$ for any $x \in \Delta$, and it is well defined even when both densities have non-common zeros. In addition, it can be expressed in terms of the relative Fisher information as

$$\overline{F}_{\rm rel}[\rho_1,\rho_2] = F_{\rm rel}\left[\rho_1, \frac{\rho_1+\rho_2}{2}\right].$$

However, it is nonsymmetric. To avoid this problem we propose the following symmetrized form

$$JFD[\rho_1,\rho_2] = \overline{F}_{\rm rel}[\rho_1,\rho_2] + \overline{F}_{\rm rel}[\rho_2,\rho_1], \tag{4}$$

as a new measure, which we call **Jensen-Fisher divergence** between the probability densities $\rho_1(x)$ and $\rho_2(x)$. From Eqs.

(4) and (3), we have that

$$
\begin{aligned}
JFD[\rho_1,\rho_2] &= F_{\text{rel}}\left[\rho_1,\frac{\rho_1+\rho_2}{2}\right]+F_{\text{rel}}\left[\rho_2,\frac{\rho_1+\rho_2}{2}\right]\\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty}\rho_1(x)\left(\frac{d}{dx}\ln\frac{\rho_1(x)}{\frac{\rho_1(x)+\rho_2(x)}{2}}\right)^2 dx\right.\\
&\qquad\left.+\int_{-\infty}^{\infty}\rho_2(x)\left(\frac{d}{dx}\ln\frac{\rho_2(x)}{\frac{\rho_1(x)+\rho_2(x)}{2}}\right)^2 dx\right)\\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty}\rho_1(x)\left(\frac{\rho_1'(x)}{\rho_1(x)}-\frac{\rho_1'(x)+\rho_2'(x)}{\rho_1(x)+\rho_2(x)}\right)^2 dx\right.\\
&\qquad\left.+\int_{-\infty}^{\infty}\rho_2(x)\left(\frac{\rho_2'(x)}{\rho_2(x)}-\frac{\rho_1'(x)+\rho_2'(x)}{\rho_1(x)+\rho_2(x)}\right)^2 dx\right)\\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty}\left(\frac{(\rho_1'(x))^2}{\rho_1(x)}-2\rho_1'(x)\frac{\rho_1'(x)+\rho_2'(x)}{\rho_1(x)+\rho_2(x)}\right.\right.\\
&\qquad\left.+\rho_1(x)\frac{(\rho_1'(x)+\rho_2'(x))^2}{(\rho_1(x)+\rho_2(x))^2}\right)dx\\
&\qquad+\int_{-\infty}^{\infty}\left(\frac{(\rho_2'(x))^2}{\rho_2(x)}-2\rho_2'(x)\frac{\rho_1'(x)+\rho_2'(x)}{\rho_1(x)+\rho_2(x)}\right.\\
&\qquad\left.\left.+\rho_2(x)\frac{(\rho_1'(x)+\rho_2'(x))^2}{(\rho_1(x)+\rho_2(x))^2}\right)dx\right)\\
&= \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{(\rho_1'(x))^2}{\rho_1(x)}+\frac{(\rho_2'(x))^2}{\rho_2(x)}-2\frac{(\rho_1'(x)+\rho_2'(x))^2}{\rho_1(x)+\rho_2(x)}\right.\\
&\qquad\left.+\frac{(\rho_1'(x)+\rho_2'(x))^2}{\rho_1(x)+\rho_2(x)}\right)\\
&= \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{(\rho_1'(x))^2}{\rho_1(x)}+\frac{(\rho_2'(x))^2}{\rho_2(x)}-\frac{(\rho_1'(x)+\rho_2'(x))^2}{\rho_1(x)+\rho_2(x)}\right),
\end{aligned}
$$

so that the Jensen-Fisher divergence can be expressed in terms of the Fisher information as

$$JFD[\rho_1,\rho_2]=\frac{F[\rho_1]+F[\rho_2]}{2}-F\left[\frac{\rho_1+\rho_2}{2}\right], \tag{5}$$

which is similar to the expression (1) of the Jensen-Shannon divergence in terms of the Shannon entropy, save for a global minus sign.

It is important to remark that the Jensen-Fisher divergence we have just introduced, shares the following properties with the Jensen-Shannon divergence. First, it is nonnegative because of Eq. (5) and the convexity of the Fisher information which leads to

$$\frac{F[\rho_1]+F[\rho_2]}{2}\geq F\left[\frac{\rho_1+\rho_2}{2}\right].$$

Second, it vanishes if and only if the two involved densities are equal almost everywhere in the interval Δ. This comes again from the fact that

$$\frac{F[\rho_1]+F[\rho_2]}{2}=F\left[\frac{\rho_1+\rho_2}{2}\right]\iff\rho_1=\rho_2,$$

by keeping in mind the convexity of the Fisher functional; so that,

$$JFD[\rho_1,\rho_2]=0\iff\rho_1=\rho_2.$$

Third, it is symmetric because one can straightforwardly prove that $JFD[\rho_1,\rho_2]=JFD[\rho_2,\rho_1]$. Fourth, it is well defined when $\rho_1(x)$ or $\rho_2(x)$ are not absolutely continuous with respect to each other (of course, as long as the Fisher information of each density is well defined) and so, it can be used to compare probability distributions with no common zeros.

In addition, the Jensen-Fisher and Jensen-Shannon divergences satisfy the following deBruijn-type expression

$$\frac{d}{d\epsilon}JSD[\rho_1+\sqrt{\epsilon}\rho_G,\rho_1+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}=-\frac{1}{2}JFD[\rho_1,\rho_2], \tag{6}$$

where ρ_G is a normal distribution with zero mean and variance equal to one. This can be proved by considering the original deBruijn's [24] identity between the Shannon entropy and the Fisher information:

$$\frac{d}{d\epsilon}S[\rho+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}=\frac{1}{2}F[\rho]. \tag{7}$$

Then,

$$
\begin{aligned}
&\frac{d}{d\epsilon}JSD[\rho_1+\sqrt{\epsilon}\rho_G,\rho_2+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}\\
&\quad=\frac{d}{d\epsilon}S\left[\frac{\rho_1+\sqrt{\epsilon}\rho_G+\rho_2+\sqrt{\epsilon}\rho_G}{2}\right]\Big|_{\epsilon=0}\\
&\quad-\frac{1}{2}\frac{d}{d\epsilon}S[\rho_1+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}-\frac{1}{2}\frac{d}{d\epsilon}S[\rho_2+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}\\
&\quad=\frac{d}{d\epsilon}S\left[\frac{\rho_1+\rho_2}{2}+\sqrt{\epsilon}\rho_G\right]\Big|_{\epsilon=0}\\
&\quad-\frac{1}{2}\frac{d}{d\epsilon}S[\rho_1+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}-\frac{1}{2}\frac{d}{d\epsilon}S[\rho_2+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}.
\end{aligned}
$$

Taking into account the deBruijn's identity (7), we obtain

$$
\begin{aligned}
&\frac{d}{d\epsilon}JSD[\rho_1+\sqrt{\epsilon}\rho_G,\rho_2+\sqrt{\epsilon}\rho_G]\Big|_{\epsilon=0}\\
&=\frac{1}{2}F\left[\frac{\rho_1+\rho_2}{2}\right]-\frac{1}{4}F[\rho_1]-\frac{1}{4}F[\rho_2]=-\frac{1}{2}JFD[\rho_1,\rho_2],
\end{aligned}
$$

and the identity (6) is proved.

Furthermore, like the Jensen-Shannon divergence, it admits a generalization to N densities with different weights $\boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_N)$, where $\omega_i\geq 0$, $i=0,1,\ldots,N$, and $\sum_{i=1}^{N}\omega_i=1$,

$$
\begin{aligned}
JFD_{\boldsymbol{\omega}}[\rho_1,\ldots,\rho_N]&=\omega_1F[\rho_1]+\cdots+\omega_NS[\rho_N]\\
&\quad-F[\omega_1\rho_1(x)+\cdots+\omega_N\rho_N(x)].
\end{aligned}
$$

Finally, let us highlight that the Jensen-Fisher divergence is informative even in those cases where the Jensen-Shannon is not. This is illustrated in Figure 1 for the particle-in-a-box system, whose stationary quantum-mechanical states are described by the probability densities (2). Therein, we have depicted the Jensen-Fisher and Jensen-Shannon divergences between the nth-state density $\rho_n(x)$ and the ground state $\rho_1(x)$, given by $JFD[\rho_n,\rho_1]$ and $JSD[\rho_n,\rho_1]$ respectively, in terms of n when n is going from 1 to 50. It turns out that, as n increases, the Jensen-Fisher divergence increases much more

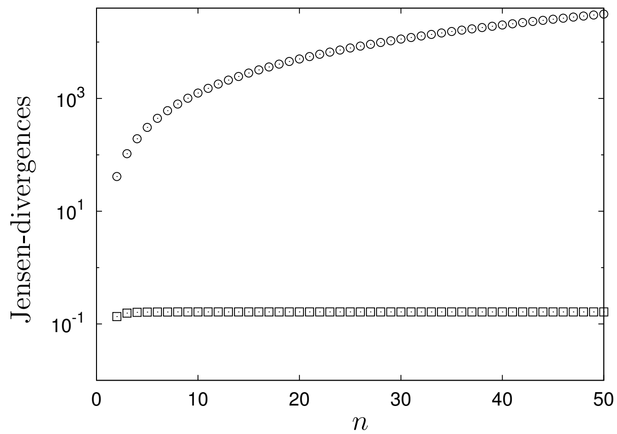

Fig. 1. Jensen-Shannon $JSD[\rho_n, \rho_1]$ ($\boxdot$) and Jensen-Fisher $JFD[\rho_n, \rho_1]$ ($\odot$) divergences between the sinusoidal densities $\rho_n(x)$ and $\rho_1(x)$ (see Eq. (2)) in terms of the quantum number n.

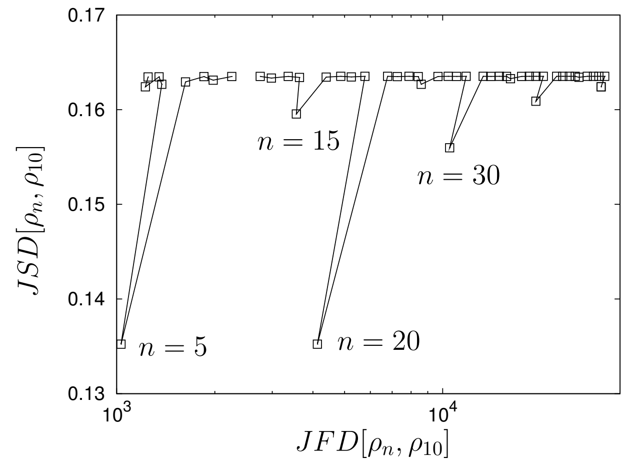

Fig. 2. $JSD[\rho_n, \rho_{10}] - JFD[\rho_n, \rho_{10}]$ divergence plane of the sinusoidal densities $\rho_n(x)$ (see Eq. (2)) for $n = 1, \ldots, 50$.

than the Jensen-Shannon, which remains practically constant. This clearly indicates that the former divergence is much more informative than the latter.

III. Jensen-Shannon and Jensen-Fisher divergences: Mutual comparison

In this Section we compare the Jensen-Fisher and the Jensen-Shannon divergences in the information-theoretic $JFD - JSD$ plane for the following three large, qualitatively different families of probability distributions: the sinusoidal distributions defined by Eq. (2), the generalized gamma-like distributions (see Eq. (8) below) and the Rakhmanov-Hermite distributions (see Eq. (9) below).

A. Sinusoidal densities

These probability densities given by Eq. (2) have been used to describe various physical systems, such as e.g. the stationary quantum-mechanical states of a particle-in-a-box (i.e., in an infinite potential well) [20] as already mentioned. Indeed, they characterized the ground state $\rho_1(x)$ and the excited states $\rho_n(x)$, with $n = 2, 3, \ldots$, of this quantum mechanical system. In Figure 1, previously discussed, we have shown that the excitation of the particle is described in a much better information-theoretical way by the JFD than by the JSD, since the former divergence between the probability density of the nth-excited-state and the ground state increases when n (so, when the energy of the particle) is increasing, while the JSD remains practically constant.

In Figure 2 we have depicted the Jensen-Shannon divergence $JSD[\rho_n, \rho_{10}]$ between the excited states with quantum number n and 10 against the corresponding Jensen-Fisher divergence $JFD[\rho_n, \rho_{10}]$ for $n = 1, \ldots, 50$. The resulting values (points) obtained for increasing n are joined by a line to guide the eye. It is observed that the JSD remains constant except for some points. They correspond to values of n multiple and submultiple of 10, that is the quantum number of the reference state ρ_{10}. At these points, $\rho_n(x)$ and $\rho_{10}(x)$ share a number of zeros, so these densities become more similar to each other, and both JSD and JFD achieve a lower value. Less dramatic deviations are observed also for values of $n = 15, 25, 35, \ldots$, where the density $\rho_n(x)$ has some of the zeros of $\rho_{10}(x)$. From a quantum-mechanical point of view, the particles on those states share some common forbidden regions (or also some common maximum probability regions). The behaviours of the JSD and JFD measures on this plane shows that although both quantities are sensitive to the overlap of the zeros, the JFD highlights this phenomenon much better (please, be aware of the different scaling in the axes of the figure). Moreover, the JFD presents larger absolute variations and has a much wider range of variation than the JSD along all the pairs of states considered.

B. Generalized gamma-like densities

In contrast with the previous case (where the densities have several zeros in a finite interval), here we consider a family of one-parameter densities having at most one zero and defined in the whole real line; namely, the gamma-like densities given by

$$\gamma_\beta(x) = \left(\sqrt{2}\, 2^{\frac{\beta}{2}} \Gamma\left(\frac{1+\beta}{2}\right)\right)^{-1} |x|^\beta \exp\left(-\frac{x^2}{2}\right); \ \beta > 1, \tag{8}$$

so that for $\beta = 0$, one has a normal distribution:

$$\gamma(x) \equiv \gamma_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

In what follows we assume that $\beta > 1$ because the Fisher information (3) is not defined for $0 < \beta \leq 1$, having a vertical asymptote at $\beta = 1$.

We have done two different analyses. First, in Figure 3, the values of $JFD[\gamma_\beta, \gamma]$ and $JSD[\gamma_\beta, \gamma]$ are given as a function of β. It shows that the Jensen-Fisher divergence is much more sensitive to the multiplicity of the zero than the Jensen-Shannon divergence. While the former varies along a range of six orders of magnitude, the latter only varies along one order of magnitude.

Second, Figure 4 shows the $JSD[\gamma, \gamma_\beta] - JFD[\gamma, \gamma_\beta]$ plane between the probability densities $\gamma(x)$ and $\gamma_\beta(x)$ for all values β from 1 to 80. Notice that there are two regimes, one for $\beta \lesssim$

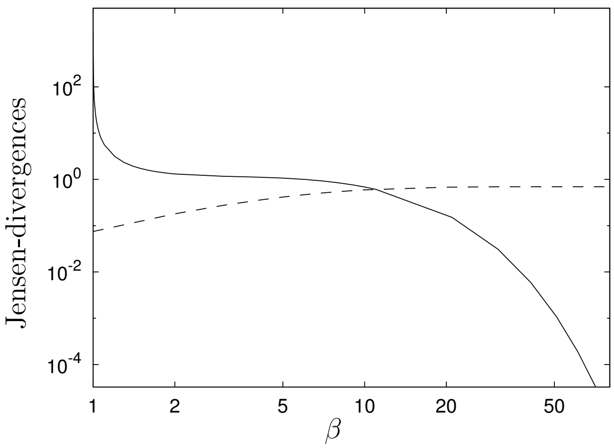

Fig. 3. Jensen-Shannon $JSD[\gamma_\beta, \gamma]$ (dashed line) and Jensen-Fisher $JFD[\gamma_\beta, \gamma]$ (solid line) divergences between the generalized gamma densities $\gamma_\beta(x)$ and $\gamma(x)$ (see Eq. (8)) as functions of the multiplicity parameter β.

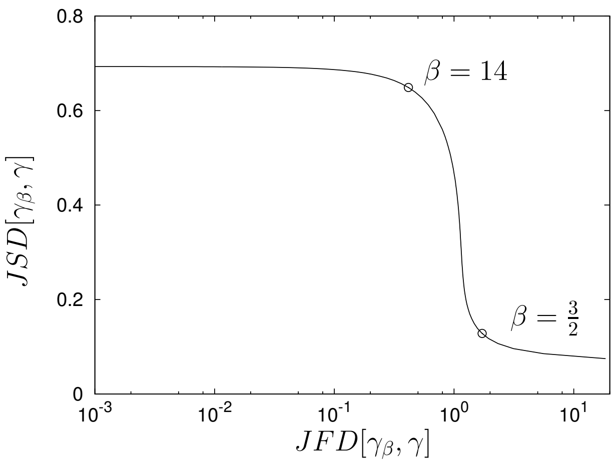

Fig. 4. $JSD[\gamma_\beta, \gamma] - JFD[\gamma_\beta, \gamma]$ divergence plane of the generalized gamma densities $\gamma_\beta(x)$ and $\gamma(x)$ (see Eq. (8)) as functions of the multiplicity parameter β.

$\frac{3}{2}$ and $\beta \gtrsim 14$ where the JSD remains almost constant and the JFD varies rapidly, and another for $\frac{3}{2} \lesssim \beta \lesssim 14$ where the JFD remains almost constant while the JSD varies. As in the previous example, the range of variation of the JFD is much wider than that of $JSD[\gamma, \gamma_\beta]$.

C. Rakhmanov-Hermite densities

Let us now consider the class of Rakhmanov-Hermite probability densities defined by

$$\rho_n^{\mathrm{HO}}(x) = \frac{1}{2^n n! \sqrt{\pi}} e^{-x^2} H_n^2(x), \tag{9}$$

where $H_n(x)$ is the orthogonal Hermite polynomial of degree n. As for the quantum infinite well previously discussed, the parameter $n = 0, 1, 2, \ldots$ indicates the energetic level and labels the corresponding state. They have been shown to correspond to the quantum-mechanical probability densities of the ground and excited stationary states of the isotropic harmonic oscillator (HO, in short); see e.g. [25], [26].

Here we have done three analyses. Firstly, we depict in Figures 5 and 6 the Jensen-Fisher and Jensen-Shannon divergences, respectively, between the nth-density and each of the reference probability densities with $n_r = 0$, 10 and 40; this is to say the quantities $JFD[\rho_n^{\mathrm{HO}}, \rho_0^{\mathrm{HO}}]$ (dotted line), $JFD[\rho_n^{\mathrm{HO}}, \rho_{10}^{\mathrm{HO}}]$ (solid line) and $JFD[\rho_n^{\mathrm{HO}}, \rho_{40}^{\mathrm{HO}}]$ (dashed line) and the corresponding JSDs. We observe from the comparison of the dotted lines of the two figures that both divergences between the nth-state density $\rho_n^{\mathrm{HO}}(x)$ and the ground-state density $\rho_0(x)$ have a increasing behaviour in terms of the quantum number n as one should expect. Moreover, from the comparison of the solid lines of the two figures, we realize an opposite behaviour in the two divergences between the nth-state density $\rho_n^{\mathrm{HO}}(x)$ and the 10th-state density $\rho_{10}^{\mathrm{HO}}(x)$ when the quantum number n (which controls the number of zeros of the density) is increasing; namely, the $JFD[\rho_n^{\mathrm{HO}}, \rho_{10}^{\mathrm{HO}}]$ has an increasing sawtooth behaviour while the $JSD[\rho_n^{\mathrm{HO}}, \rho_{10}^{\mathrm{HO}}]$ firstly decreases down to zero when n goes from 0 to the reference number 10, and then increases when n goes from 10 upwards. A similar trend is observed from the comparison of the dashed lines of the two figures for the $JFD[\rho_n^{\mathrm{HO}}, \rho_{40}^{\mathrm{HO}}]$ and $JSD[\rho_n^{\mathrm{HO}}, \rho_{40}^{\mathrm{HO}}]$ divergences but now with respect to the reference number 40. Clearly, in the three cases $(n, n_r) = (n, 1)$, $(n, 10)$ and $(n, 40)$ the Jensen-Fisher divergence has always higher variations than the Jensen-Shannon divergence, because the JFD has a stronger sensitivity than the JSD to the increasing oscillatory character of $\rho_n(x)$ when n is increasing. In addition we observe that the Jensen-Fisher divergence presents two maxima around the reference value (maxima at 9 and 11 for $n = 10$, and at 39 and 41 for $n = 40$) that can be explained taking into account the relative position of the zeros of the two involved densities. In those cases each zero of one of the densities is situated between two zeros of the other density, so none of the zeros of one of the densities are near the zeros of the other one. The opposite situation occurs for the local minima that appears in the graphics, where some zeros of a density are near the zeros of the other one. The Jensen-Shannon divergence also shows the latter feature but with much less intensity. However, it does not show the local maxima around the reference value.

Our second analysis is shown in Figure 7, where we study the comparison of the JSD and JFD divergences between the pairs of probability densities with quantum numbers $(n, 0)$, $(n, 10)$ and $(n, 40)$ in the frame of the $JSD - JFD$ divergence plane. This figure combines the results contained in the two previous Figures 5 and 6, and shows again the overall increasing behaviour of the Jensen-Fisher divergence and its much higher values, in contrast to the Jensen-Shannon divergence. The most important feature that this Figure shows is the separation of the clouds of points in the direction of increasing JFD, while these clouds are not distinguishable from their JSD values. Let us mention that the two couples of points to the right of the vertices of the V-shaped structures, correspond to the local maxima that appear in Figure 5.

Finally, in Figure 8 we use again the $JSD - JFD$ plane as a tool to simultaneously show the distance or divergence of the pairs of probability densities with quantum numbers $(n, n+1)$, $(n, n+10)$, $(n, 2n)$, $(n, 2n+10)$, $(n, 3n)$ and

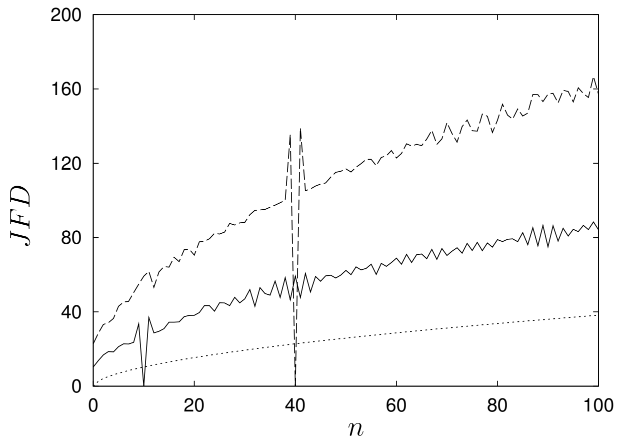

Fig. 5. Jensen-Fisher divergences $JFD[\rho_n^{\mathrm{HO}}, \rho_0^{\mathrm{HO}}]$ (dotted line), $JFD[\rho_n^{\mathrm{HO}}, \rho_{10}^{\mathrm{HO}}]$ (solid line) and $JFD[\rho_n^{\mathrm{HO}}, \rho_{40}^{\mathrm{HO}}]$ (dashed line) between the nth-excited state and the ground state, $10th$ and $40th$-excited states of the isotropic harmonic oscillator, respectively, in terms of the quantum number n.

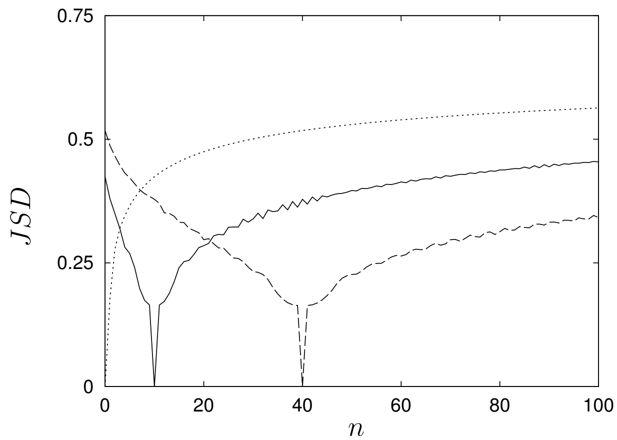

Fig. 6. Jensen-Shannon divergences $JSD[\rho_n^{\mathrm{HO}}, \rho_0^{\mathrm{HO}}]$ (dotted line), $JSD[\rho_n^{\mathrm{HO}}, \rho_{10}^{\mathrm{HO}}]$ (solid line) and $JSD[\rho_n^{\mathrm{HO}}, \rho_{40}^{\mathrm{HO}}]$ (dashed line) between the nth-excited state and the ground state, $10th$ and $40th$-excited states of the isotropic harmonic oscillator, respectively, in terms of the quantum number n.

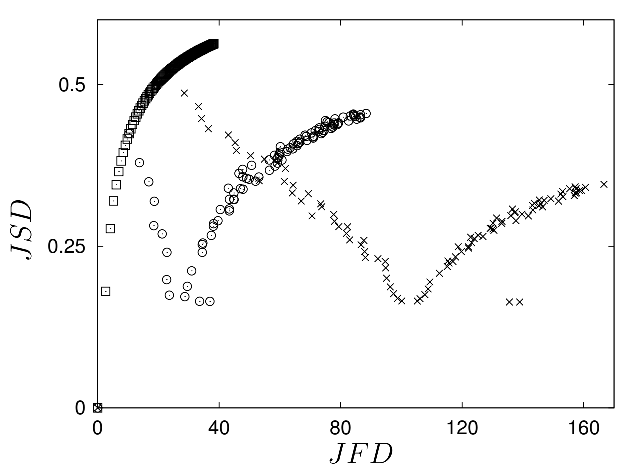

Fig. 7. $JSD - JFD$ divergence plane of the isotropic harmonic oscillator for the pairs of stationary states $(n, m) = (n, 0)$ ($\boxdot$), $(n, 10)$ ($\odot$) and $(n, 40)$ ($\times$) when n varies from 0 to 100.

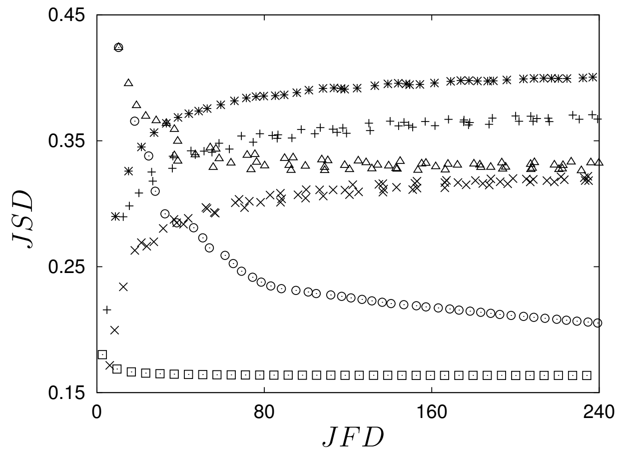

Fig. 8. $JSD - JFD$ divergence plane of the isotropic harmonic oscillator for the pairs of stationary states $(n, m) = (n, n+1)$ ($\boxdot$), $(n, n+10)$ ($\odot$), $(n, 2n)$ ($\times$), $(n, 2n{+}10)$ ($\triangle$), $(n, 3n)$ ($+$) and $(n, 4n)$ ($*$), for several values of n from $n = 0$ up to a value of the JFD of 240.

$(n, 4n)$. Here we notice that, as n increases, the JFD tends to infinity in all the cases, but the JSD tends to a constant. We observe that $JSD[\rho_n^{\mathrm{HO}}, \rho_{n+1}^{\mathrm{HO}}]$ tends to the same value as $JSD[\rho_n^{\mathrm{HO}}, \rho_{n+10}^{\mathrm{HO}}]$, and $JSD[\rho_n^{\mathrm{HO}}, \rho_{2n}^{\mathrm{HO}}]$ tends to the same value as $JSD[\rho_n^{\mathrm{HO}}, \rho_{2n+10}^{\mathrm{HO}}]$, being those two limiting values different from each other. Then, we can conclude that this asymptotic value of the JSD depends on the relative spreading of the involved densities. When n tends to infinity, the spreading of the density $\rho_n^{\mathrm{HO}}(x)$ converges to that of $\rho_{n+1}^{\mathrm{HO}}(x)$ or $\rho_{n+10}^{\mathrm{HO}}(x)$. However, $\rho_n^{\mathrm{HO}}(x)$ is less spread than $\rho_{2n}^{\mathrm{HO}}(x)$ or $\rho_{2n+10}^{\mathrm{HO}}(x)$. Thus, the JSD between those densities tends to a different value. This trend is confirmed by the asymptotic values of $JSD[\rho_n^{\mathrm{HO}}, \rho_{3n}^{\mathrm{HO}}]$ and $JSD[\rho_n^{\mathrm{HO}}, \rho_{4n}^{\mathrm{HO}}]$.

As in previous analyses, Figure 8 shows that the JFD has a much wider range of variation than the JSD, so that it allows us to discriminate between different values of n in a better way. However, contrary to what happened in Figure 7, the JFD cannot distinguish between the different clouds of points of Figure 8. This is a clear illustration of the complementarity of both the JSD and JFD when analysing the similarity of probability distributions.

IV. Conclusions and open problems

In this paper the Jensen-Fisher divergence measure is introduced and its theoretical grounds are shown. In summary, we find that the main properties (non-negativity, additivity, symmetry, vanishing, definiteness, deBruijn-like identity) of the Jensen-Shannon divergence are shared by the new divergence. Moreover, the Jensen-Fisher divergence is applied to three large families of representative probability distributions (sinusoidal, gamma-like and Rakhmanov-Hermite distributions) and compared with the Jensen-Shannon divergence. Our results illustrate that, although both JSD and JFD divergences are complementary in the sense that they are sensitive to different aspects of the probability distributions, the latter is more informative when studying the similarity of oscillating densities.

Finally we should immediately point out three open issues. First, the square root of JSD is known to define a metric [5], [27]. Does the Jensen-Fisher divergence defines another distance metric for probability distributions beyond the JSD [5]–[7], [27] and the variational distance [7]?. This is still an open problem which deserves much attention *per se* and because of its so many implications in numerous scientific and technological fields. Second, some generalizations of the JSD have been recently introduced such as the Jensen-Rényi [28] and Jensen-Tsallis [18], [29], [30] divergences as well as the Jensen divergences of order α [31] paying the price of the loss of certain interesting properties but gaining more flexibility because they have a new degree of freedom provided by its parameter q or α, what is very useful in numerous applications (see e.g., [32]–[36]). Does the Jensen-Fisher divergence admits any generalization?. The answer is yes but this avenue is still to be paved. Finally, does there exist a quantum version of the JFD based on the quantum Fisher information [37], [38] similarly to the quantum JSD based on the von Neuman entropy [27], [31], [39]–[42]?

Acknowledgement

PSM and JSD are very grateful to Junta de Andalucía for the grants FQM-2445 and FQM-4643, and the Ministerio de Ciencia e Innovacíon for the grant FIS2008-02380. PSM and JSD belong to the research group FQM-207.

AZ agknowledges partial fiancial support from Ministerio de Educación y Ciencia of Spain under grants MTM2006-07186 and MTM2009-14668-C02-02 and from Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía, Spain, under grant P09-TEP-5022. Also, AZ has been partially funded by UPM under some contracts.

This work was finished while on a staying of AZ at Instituo Carlos I of the Granada University partly funded by this Institute and also by the Departamento de Matemática Aplicada a la Ingeniería Industrial, ETSII, UPM.

References

[1] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Information Theory*, vol. 37, pp. 145–151, 1991.

[2] C. R. Rao, "Differential Geometry in Statistical Inference," *IMS-Lecture Notes*, vol. 10, pp. 217–225, 1987.

[3] S. Kullback and A. Leibler, "On the information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 1951.

[4] S. Kullback, *Information Theory and Statistics*. Dover Publications, New York, 1968.

[5] D. M. Endress and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Information Theory*, vol. 49, pp. 1858–1860, 2003.

[6] F. Topsøe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans. Information Theory*, vol. 46, pp. 1602–1609, 2000.

[7] S. C. Tsai, W. G. Tzeng, and H. L. Wu, "On the Jensen-Shannon divergence and variational distance," *IEEE Trans. Information Theory*, vol. 51, pp. 3333–3336, 2005.

[8] P. W. Lamberti, A. P. Majtey, M. Madrid, and M. Pereyra, in *Proceed. of the XV Conference on Non-equilibrium Statistical Mechanics and Nonlinear Physics*, O. Descalzi, O. A. Rosso, and H. A. Larrondo, Eds. American Institute of Physics, New York, 2007, pp. 32–37.

[9] R. Román-Roldán, P. Bernaola-Galván, and J. Oliver, "Sequence compositional complexity of DNA through an entropic segmentation method," *Phys. Rev. Lett.*, vol. 80, pp. 1344–1347, 1998.

[10] G. E. Simms, S. R. Jun, G. A. Wu, and S. H. Kim, "Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions," *Proc. Natl. Acad. Sci. USA*, vol. 106, pp. 2677–2682, 2009.

[11] J. C. Angulo, J. Antolin, S. López-Rosa, and R. O. Esquivel, "Jensen-Shannon divergence in conjugate spaces: The entropy excess of atomic systems and sets with respect to their constituents," *Physica A*, vol. 389, p. 899, 2010.

[12] J. Antolin, J. C. Angulo, and S. López-Rosa, "Fisher and Jensen-Shannon divergences: Quantitative comparisons among distributions. Application to position and momentum atomic densities," *J. Chem. Phys.*, vol. 130, p. 074110, 2009.

[13] K. C. Chatzisavvas, C. C. Moustakidis, and C. Panos, "Information entropy, information distances, and complexity in atoms," *J. Chem. Phys.*, vol. 123, p. 174111, 2005.

[14] S. López-Rosa, J. Antolín, J. C. Angulo, and R. O. Esquivel, "Divergence analysis of atomic ionization processes and isoelectronic series," *Phys. Rev. A*, vol. 80, p. 012505, 2009.

[15] A. Majtey, P. W. Lamberti, M. T. Martin, and A. Plastino, "Wootters' distance revisited: a new distinguishability criterium," *Eur. Phys. J. D*, vol. 32, pp. 413–419, 2005.

[16] A. P. Majtey, A. Borras, A. R. Plastino, M. Casas, and A. Plastino, "Some feature of the state-space trajectories followed by robust entangled four-qubit states during decoherence," *Int. J. Quantum Inf.*, vol. 8, pp. 505–515, 2010.

[17] M. E. Pereyra, P. W. Lamberti, and O. Rosso, "Wavelet Jensen-Shannon divergence as a tool for studying the dynamics of frequency band components in EEG epileptic seizures," *Physica A*, vol. 379, pp. 122–132, 2007.

[18] P. W. Lamberti and A. P. Majtey, "Non-logarithmic Jensen-Shannon divergence," *Physica A*, vol. 329, pp. 81–86, 2003.

[19] D. H. Zanette, "Segmentation and context of literary and musical sequences," *Complex Systems*, vol. 17, pp. 279–293, 2007.

[20] A. Galindo and P. Pascual, *Quantum Mechanics*. Springer, Berlin, 1990.

[21] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Phil. Soc.*, vol. 22, pp. 700–725, 1925, reprinted in Collected Papers of R.A. Fisher, edited by J.H. Bennet (University of Adelaide Press, South Australia), 1972, 15–40.

[22] B. R. Frieden, *Science from Fisher Information*. Cambridge University Press, Cambridge, 2004.

[23] P. Hammad, "Mesure d'ordre α de l'information au sens de Fisher," *Revue de Statistique Appliquée*, vol. 26, pp. 73–84, 1978.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, N.Y., 1991.

[25] A. F. Nikiforov and V. B. Uvarov, *Special Functions in Mathematical Physics*. Birkäuser-Verlag, Basel, 1988.

[26] P. Sánchez-Moreno, J. S. Dehesa, D. Manzano, and R. J. Yáñez, "Spreading lengths of Hermite polynomials," *J. Comput. Appl. Math.*, vol. 233, pp. 2136–2148, 2010.

[27] P. W. Lamberti, A. P. Majtey, A. Borras, M. Casas, and A. Plastino, "Metric character of the quantum Jensen-Shannon divergence," *Phys. Rev. A*, vol. 77, p. 052311, 2008.

[28] J. Burbao and C. R. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Trans. Information Theory*, vol. 28, p. 489, 1982.

[29] A. B. Hamza, "Nonextensive information-theoretic measure for image edge detection," *J. Electron. Imaging*, vol. 15, p. 013011, 2006.

[30] A. P. Majtey, P. W. Lamberti, and A. Plastino, "A monoparametric family of metrics for statistical mechanics," *Physica A*, vol. 344, pp. 547–553, 2004.

[31] J. Briët and P. Harremoës, "Properties of classical and quantum Jensen-Shannon divergence," *Phys. Rev. A*, vol. 79, p. 052311, 2009.

[32] A. B. Hamza and H. Krim, "Jensen-Rényi divergence measure: theoretical and computational perspectives," in *IEEE International Symposium on Information Theory ISIT*, 2003, p. 257.

[33] Y. He, A. B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Trans. Signal Proc.*, vol. 51, p. 1211, 2003.

[34] M. C. Chiang, R. A. Dutton, K. M. Hayashi, A. W. Toga, O. L. Lopez, H. J. Aizenstein, J. T. Becker, and P. M. Thompson, "Fluid registration of medical images using Jensen-Rényi divergence reveals 3D profile of brain atrophy in HIV/AIDS," in *International Symposium on Biomedical Imaging*, 2006, p. 193.

[35] F. Wang, T. Syeda-Mahmood, B. C. Vemuri, D. Beymer, and A. Rangarajan, "Closed-form Jensen-Renyi divergence for mixture of Gaussians and applications to group-wise shape registration," in *Lecture Notes in Computer Scienc., Vol. 5761*, 2009, pp. 648–655.

[36] J. Antolin, S. López-Rosa, J. C. Angulo, and R. O. Esquivel, “Jensen-Tsallis divergence and atomic dissimilarity for position and momentum space electron densities,” *J. Chem. Phys.*, vol. 132, p. 044105, 2010.

[37] S. Luo, “Quantum Fisher information and uncertainty relations,” *Lett. Math. Phys.*, vol. 53, pp. 243–251, 2000.

[38] P. Gibilisco, F. Hiai, and D. Petz, “Quantum covariance, quantum Fisher information, and the uncertainty relations,” *IEEE Trans. Information Theory*, vol. 55, pp. 439–443, 2009.

[39] A. P. Majtey, P. W. Lamberti, and D. P. Prato, “Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states,” *Phys. Rev. A*, vol. 72, p. 052310, 2005.

[40] P. W. Lamberti, M. Portesi, and J. Sparacino, “Natural metric for quantum information theory,” 2009, arXiv:0807.0583v2[quant-ph].

[41] S. L. Braunstein and C. M. Caves, “Statistical distance and the geometry of quantum states,” *Phys. Rev. Lett.*, vol. 72, p. 3439, 1994.

[42] W. Roga, M. Fannes, and K. Zyczkowski, “Universal bounds fo the Holevo quantity, coherent information and the Jensen-Shannon divergence,” *Phys. Rev. Lett.*, vol. 105, p. 040505, 2010.